A20 concl'd

with real-time word and document updates, whereas SVD re-computation would generally not be feasible.

REMARKS

Please charge any shortages and credit any overcharges to our Deposit Account number 02-2666.

Respectfully submitted,

BLAKELY, SOKOLOFF, TAYLOR & ZAFMAN, LLP

Date: 12/27/2001

Donna Jo Coningsby
Reg. No. 41,684

12400 Wilshire Boulevard
Seventh Floor
Los Angeles, CA 90025-1026
(503) 684-6200

OIPE JC185
FEB 1 2 2002
PATENT & TRADEMARK OFFICE

APPENDIX A:

VERSION WITH MARKINGS TO SHOW CHANGES MADE TO SPECIFICATION

The paragraph beginning on page 6 at line 10:

FIGS. 3A and 3B [is] are an overview of selected components of the basic LSA paradigm, in accordance with one embodiment of the present invention;

The paragraph beginning on page 6 at line 12:

FIGS. 4A and 4B [is] are an overview of selected components of the adaptive LSA paradigm, in accordance with one embodiment of the present inventions;

The paragraph beginning on page 6 at line 14:

FIGS. 5A and 5B [is] are an overview of selected components of the matrix transformation of the adaptive LSA paradigm, in accordance with one embodiment of the present invention;

The paragraph beginning on page 6 at line 16:

FIGS. 6A and 6B [is] are an overview of selected components of the vector transformation of the adaptive LSA paradigm, in accordance with one embodiment of the present invention;

The paragraph beginning on page 6 at line 18:

FIGS. 7A and 7B [is] are an overview of selected components of prior art baseline adaptation;

The paragraph beginning on page 11 at line 18:

FIGS. 3A and 3B illustrate[s] selected components of the basic LSA paradigm 300 used to construct the continuous vector space $\mathcal{S}$, referenced in FIG. 3B as LSA space

$\mathcal{S}$ 316. The LSA paradigm 300 first captures the semantic patterns of the word-document co-occurrences that appeared in the training corpus $\mathcal{T}$ 202 by constructing a word-document matrix W 302 of dimension $M \times N$, whose entries w_{ij} 304 suitably reflect the extent to which word w_i 208 appeared in document d_j 204, and then performing a singular value decomposition (SVD) of the word-document matrix W 302 having an order of decomposition of $R << \min(M,N)$ as in [1]:

$$W = USV^T, \tag{1}$$

where U 306 is the $M \times R$ left singular matrix of row vectors, $u_i(1 \le i \le M)$, S 308 is the $R \times R$ diagonal matrix of singular values $s_1 \ge s_2 \ge \ldots s_R > 0$, and V^T is the transposition of V 310, the $R \times N$ right singular matrix of row vectors $v_j(1 \le j \le N)$. The value of R can vary depending on the values of M and N, and by balancing computational speed (associated with lower values of R) against accuracy (associated with higher values of R). Typical values for R range from 5 to 100.

The paragraph beginning on page 13 at line 9:

FIGS. 4A and 4B illustrate[s] selected components of the adaptive LSA paradigm 400 using latent semantic adaptation in accordance with an embodiment of the present invention. The adaptive LSA paradigm 400 extends the basic LSA paradigm 300 so that some or all of the data in new documents 110 are taken into account through incremental adaptation of the original LSA space $\mathcal{S}$ 316 in a way that is computationally efficient. Adaptation of the original LSA space $\mathcal{S}$ 316 insures that the semantic classification error rate of the semantic classification unit 112 does not substantially increase as the new

words and documents 110 vary from those contained in the original training corpus 7 202.

The paragraph beginning on page 13 at line 25:

With reference to FIG. 4A, if n additional documents contain words drawn from the original underlying vocabulary V 206 plus m words previously unseen (i.e. out-of-vocabulary words), then the adaptive LSA paradigm 400 constructs a word-document matrix $\widetilde{W}$ 402 of dimension $(M+m)\times(N+n)$ in the same manner as described for generating matrix W 202 in the basic LSA paradigm 300 in FIGS. 3A and 3B. Using the same order of decomposition R, the SVD of $\widetilde{W}$ 402 leads to:

$$\widetilde{W} = \widetilde{U}\widetilde{S}\widetilde{V}^T, \tag{4}$$

where $\widetilde{U}$ 406 is the left singular matrix of dimension $(M+m)\times R$, $\widetilde{S}$ 408 is the diagonal matrix of dimension $R \times R$, and $\widetilde{V}$ 410 is the right singular matrix of dimension $(N+n)\times R$, each having the same definitions and properties as described above for W, U, S, and V in FIGS. 3A and 3B.

The paragraph beginning on page 14 at line 9:

As shown in FIG. 4A, the m new words are gathered in the $m\times(N+n)$ matrix $\widetilde{C} = [CE]$ 422, the n new documents are gathered in the $(M+m)\times n$ matrix $\widetilde{D} = \left[D^T E^T\right]^T$ 424. $\widetilde{U}$ 406 is expressed as $\left[\widetilde{U}_1^T \widetilde{U}_2^T\right]^T$, where $\widetilde{U}_1^T$ 436 is the transposition of the left singular matrix of dimension $M \times R$ and $\widetilde{U}_2^T$ 438 is the transposition of the left singular matrix of dimension $m \times R$. $\widetilde{V}^T$ 410 is expressed as $\left[\widetilde{V}_1^T \widetilde{V}_2^T\right]$ where $\widetilde{V}_1^T$ 439 is the transposition of the right singular matrix of dimension $R \times N$ and $\widetilde{V}_2^T$ 440 is the

transposition of the right singular matrix of dimension $R \times n$. The new decomposition of $\tilde{W}$ expressed in (4) leads to a different LSA space $\tilde{\mathcal{S}}$ 416, in which the word and document vectors are now given by the scaled row vectors $\bar{\tilde{u}}_i = \tilde{u}_i \tilde{S}$ 418 and $\bar{\tilde{v}}_j = \tilde{v}_j \tilde{S}$ 420 (i.e. the rows of $\widetilde{US}$ 412 and $\widetilde{VS}$ 414) to characterize the position of word w_i and document d_j.

The paragraph beginning on page 14 at line 20:

FIGS. 7A and 7B illustrate[s] the prior art approach referred to as baseline adaptation 700, where the distinction between the SVD in (1) of the original word-document co-occurrence matrix W 302 in FIG. 3A and the SVD in (4) of the extended word-document co-occurrence matrix $\tilde{W}$ 402 in FIG. 4A is ignored by making the (obviously invalid) assumption that the original LSA space $\mathcal{S}$ 316 in FIG. 3B is the same as the new LSA space $\tilde{\mathcal{S}}$ 416 in FIG. 4B. In other words, in baseline adaptation 700, the SVD in (1) is still assumed to be valid even after the new documents become available, and the problem is reduced to representing the new data in the original LSA space $\mathcal{S}$ 316.

The paragraph beginning on page 15 at line 1:

Referring now to FIGS. 4A and 7B, the baseline adaptation approach 700 treats the portions of the matrix $\tilde{W}$ 402 identified as C 430 and D 432 as merely extensions of additional rows or columns of the original matrix W 302, and discards altogether the portion of the extended matrix $\tilde{W}$ 402 identified as E 434. This has the effect of ignoring significant amounts of new data, including any out-of-vocabulary words in the new documents.

The paragraph beginning on page 15 at line 7:

Using the baseline adaptation approach <u>700</u>, the representation of those portions of the new data that will be added to the original LSA space $\mathcal{S}$ 316 is obtained from the SVD of as C 430 and D 432 as follows:

$$C = YSV^T, \tag{5}$$

$$D = USZ^T, \tag{6}$$

where the $m \times R$ matrix Y 426 and the $n \times R$ matrix Z 428 are defined *a posteriori* (as plug-ins), to satisfy the relationship. In essence, using the baseline adaptation framework 700, the role of matrices Y 426 and Z 428 is to "extend" the original matrices U 306 and V 310 to accommodate the new data. The original word and document vectors $\bar{u}_i$ 318 and $\bar{v}_j$ 320 are still given by the rows of US 312 and VS 314, but the new word and document vectors $\bar{y}_i$ 446 and $\bar{z}_j$ 448 are given by the rows of YS 442 and ZS 444, respectively. From (5) and (6), these are seen to be:

$$YS = CV, \tag{7}$$
$$ZS = D^T U. \tag{8}$$

The effect, illustrated in FIG. 7<u>B</u>, is that the original LSA space $\mathcal{S}$ 316 becomes populated with the new data, i.e. the new word and document vectors $\bar{y}_i$ 446 and $\bar{z}_j$ 448, hence the name "folding-in."

The paragraph beginning on page 15 at line 24:

A major drawback to the above-described baseline adaptation approach <u>700</u> illustrated in FIG. 7<u>B</u> is poor performance, since even when populated with the new word and document vectors $\bar{y}_i$ 446 and $\bar{z}_j$ 448, the misclassification error rate using the original LSA space $\mathcal{S}$ 316 is still high when the new words and documents vary from the

original training corpus 7202, e.g. when the new documents contain several new words not in the original training corpus.

The paragraph beginning on page 16 at line 3:

In contrast, the latent semantic adaptation approach of the present invention achieves significant reductions in the misclassification error rate. Unlike baseline adaptation 700, the latent semantic adaptation approach of the present invention recognizes that there is an important distinction between the SVD in (1) of the original word-document co-occurrence matrix W 302 in FIG. 3A and the SVD in (4) of the extended word-document co-occurrence matrix $\tilde{W}$ 402 in FIG. 4A that must be taken into account since the original LSA space $\mathcal{S}$ 316 in FIG. 3B is not the same as the new LSA space $\tilde{\mathcal{S}}$ 416 in FIG. 4B. In other words, the SVD in (1) is no longer valid after the new documents become available, so the problem is more than just representing the new data in the original LSA space $\mathcal{S}$ 316. Therefore, in one embodiment, the latent semantic adaptation approach treats the portions of the matrix $\tilde{W}$ 402 identified as C 430 and/or D 432 in FIG. 4A as new data that must be accounted for in a new LSA space $\tilde{\mathcal{S}}$ 416. In one embodiment, the portion of the matrix $\tilde{W}$ 402 identified as E 434 in FIG. 4A is also treated as new data that must be accounted for in a new LSA space $\tilde{\mathcal{S}}$ 416.

The paragraph beginning on page 16 at line 17:

In one embodiment of latent semantic adaptation, the scaled row vectors (i.e. the rows of $\tilde{U}\tilde{S}$ 412 and $\tilde{V}\tilde{S}$ 414) are obtained directly from the SVD of the entire matrix

$\widetilde{W}$ 402 in (4) using a latent semantic adaptation framework 400 as defined in the equations that follow. By inspection from FIG. 4A,

$$C = \widetilde{U}_2 \widetilde{S} \widetilde{V}_1^T, \tag{9}$$

$$D = \widetilde{U}_1 \widetilde{S} \widetilde{V}_2^T, \tag{10}$$

and

$$W = \widetilde{U}_1 \widetilde{S} \widetilde{V}_1^T, \tag{11}$$

$$E = \widetilde{U}_2 \widetilde{S} \widetilde{V}_2^T, \tag{12}$$

each of which are column-orthonormal, i.e., $\widetilde{U}^T \widetilde{U} = \widetilde{V}^T \widetilde{V} = I_R$ (the identity matrix of order R). The orthogonality constraints can also be expressed in terms of $\widetilde{U}_1$, $\widetilde{U}_2$, $\widetilde{V}_1$, and $\widetilde{V}_2$ as follows:

$$\widetilde{U}^T \widetilde{U} = I_R = \widetilde{U}_1^T \widetilde{U}_1 + \widetilde{U}_2^T \widetilde{U}_2, \tag{13}$$

$$\widetilde{V}^T \widetilde{V} = I_R = \widetilde{V}_1^T \widetilde{V}_1 + \widetilde{V}_2^T \widetilde{V}_2. \tag{14}$$

In one embodiment, the foregoing equations (9)-(14) define the latent semantic adaptation framework 400 of the method of the present invention. The latent semantic adaptation framework 400 is used to solve for the "extension" SVD matrices $\widetilde{U}$ 406, $\widetilde{S}$ 408, and $\widetilde{V}$ 410 as a function of the original SVD matrices U 306, S 308, V 310, and "extension" SVD matrices Y 426, and Z 428.

The paragraph beginning on page 17 at line 11:

According to one embodiment, the solution is obtained by setting up a latent semantic adaptation transformation 500, as illustrated in FIGS. 5A and 5B, based on the assumptions previously noted that the dimension R of the original LSA space $\mathcal{S}$ 316 is low enough that none of the corresponding R singular values are zero, and that the

transformation necessary to adapt the original LSA space $\mathcal{S}$ 316 is invertible. Starting with $\widetilde{S}$ 408, the shift from S 308 in FIG. 3A to $\widetilde{S}$ 408 in FIG. 4A can be captured as illustrated in FIGS. 5A and 5B by the following expressions:

$$\widetilde{U}_1 = U\widetilde{G}, \tag{15}$$
$$\widetilde{V}_1 = V\widetilde{H}, \tag{16}$$

where $\widetilde{G}$ 508 and $\widetilde{H}$ 518 are (R x R) matrices that, according to the second assumption, are assumed to be invertible. Taken together, (15) and (16) define a latent semantic adaptation matrix transformation 500 to apply to the original SVD matrices U 306 and V 310 to update them according to the new data.

The paragraph beginning on page 20 at line 7:

From equations (17), (28), and (29), it is clear that:

$$(\widetilde{G}\widetilde{S})(\widetilde{G}\widetilde{S})^T = \widetilde{G}\widetilde{S}^2\widetilde{G}^T = S\widetilde{H}^{-T}\widetilde{H}^{-1}S = S(I_R + Z^T Z)S, \tag{37}$$
$$(\widetilde{H}\widetilde{S})(\widetilde{H}\widetilde{S})^T = \widetilde{H}\widetilde{S}^2\widetilde{H}^T = S\widetilde{G}^{-T}\widetilde{G}^{-1}S = S(I_R + Y^T Y)S. \tag{38}$$

Thus, it is also possible to obtain $\widetilde{G}\widetilde{S}$ and $\widetilde{H}\widetilde{S}$ directly through Choleski decomposition, in a manner analogous to that mentioned above $\widetilde{G}$ 508 and $\widetilde{H}$ 518. In fact, as illustrated in FIGS. 6A and 6B, if J [608] 618 and K [618] 608 are the solutions of relevant Choleski decompositions, viz.:

$$JJ^T = (I_R + Y^T Y), \tag{39}$$
$$KK^T = (I_R + Z^T Z), \tag{40}$$

then equations (35)-(38) admit as solutions:

$$\widetilde{U}\widetilde{S} = \begin{bmatrix} US \\ YS \end{bmatrix} K, \tag{41}$$
$$\widetilde{V}\widetilde{S} = \begin{bmatrix} VS \\ ZS \end{bmatrix} J. \tag{42}$$

The paragraph beginning on page 20 at line 20:

In other words, in accordance with one embodiment of the present invention, the original vectors US 312 and VS 314, as well as the new vectors resulting from the "folding-in" process YS 442 and ZS 444, can be transformed using a latent semantic adaptation vector transformation 600 defined by the transformation matrices K [618] <u>608</u> in FIG. 6<u>A</u> and J [608] <u>618</u> in FIG 6<u>B</u> to respectively yield the updated word vectors $\widetilde{US}$ 412 and document vectors $\widetilde{VS}$ 414. Therefore, equations (41) and (42) make it possible to adapt the original LSA space $\mathcal{S}$ 316 of FIG. 3<u>B</u> to the new LSA space $\tilde{\mathcal{S}}$ 416 of FIG. 4<u>B</u>.

The paragraph beginning on page 21 at line 3:

In one embodiment of the latent semantic adaptation framework 400, the new information, as reflected through the transformation matrices K [618] <u>608</u> and J [608] <u>618</u>, affects both original word and document vectors $\bar{u}_i$ 318 and $\bar{v}_j$ 320 and new word and document vectors $\bar{y}_i$ 446 and $\bar{z}_j$ 448, referred to as two-sided adaptation. Stated another way, the transformed representation of the new word and document vectors $\bar{y}_i$ 446 and $\bar{z}_j$ 448 takes into account its own influence on the underlying semantic knowledge that was encapsulated in the original LSA space $\mathcal{S}$ 316 of FIG. 3<u>B</u> (i.e. the existing word and document vectors $\bar{u}_i$ 318 and $\bar{v}_j$ 320) to yield the transformed word and document vectors $\bar{\tilde{u}}_i$ 418 and $\bar{\tilde{v}}_j$ 420 that populate the new LSA space $\tilde{\mathcal{S}}$ 416 of FIG. 4<u>B</u>. As indicated by the arrows in the new LSA space $\tilde{\mathcal{S}}$ 416 of FIG. 4<u>B</u>, the

positions of both the words and documents represented by original word and document vectors $\bar{u}_i$ 318 and $\bar{v}_j$ 320 have shifted from their positions in the original LSA space $\mathcal{S}$ 316 to reflect their changed position (i.e. their relationship) within the new LSA space $\tilde{\mathcal{S}}$ 416. The new LSA space $\tilde{\mathcal{S}}$ 416 allows not only for improvements in the misclassification error rate, but also provides the ability to adapt the speech recognition database that embodies the new LSA space $\tilde{\mathcal{S}}$ 416 in real-time, because the application of the transformation matrices K [618] <u>608</u> and J [608] <u>618</u> is computationally efficient and bypasses the need to re-compute the LSA space.

The paragraph beginning on page 22 at line 3:

In addition to providing improved performance through lowering the misclassification rate, it is also worth noting that the latent semantic adaptation framework 400 and resulting latent semantic adaptation matrix and vector transformations 500 and 600 respectively are computationally efficient. Compared to the "folding-in" computations of the baseline adaptation approach <u>700</u>, the latent semantic adaptation matrix and vector transformations 500 and 600 of the latent semantic adaptation framework 400 entail less overhead. For example, in terms of the number of floating point operations required, the overhead associated with the latent semantic adaptation vector transformations 600 embodied in equations (39)-(42) can be expressed as:

$$N_{adapt} = \frac{2}{3}R^3 + \left[(M+N) + 2(m+n) - 1\right]R^2 + (m+n+1)R. \qquad (43)$$

For typical values of the various dimensions involved, expression (43) will be dominated by $(M+N)R^2$. Depending on the application, this quantity may fall anywhere between about 50 million (for voice command and control types of speech recognition applications using a limited vocabulary) and more than 1 billion (for large vocabulary transcription). Still, on current high-end machines, this quantity only represents up to a few seconds of central processor unit (CPU) time. Compared to recomputing the SVD from scratch, which requires $O(MNR)$ operations, the computational complexity is reduced by a factor of approximately $\min(M,N)/R$. In many speech recognition applications, the reduction factor will be on the order of 1000. In such cases, the latent semantic adaptation framework 400 and resulting latent semantic adaptation matrix and vector transformations 500 and 600 make it practical to adapt the new LSA space $\tilde{S}$ 416 with real-time word and document updates, whereas SVD re-computation would generally not be feasible.